

# Using Real Data in an SIR Model

D. Sulsky

June 21, 2012

In most epidemics it is difficult to determine how many new infectives there are each day since only those that are removed, for medical aid or other reasons, can be counted. Public health records generally give the number of removed per day, per week, or per month. So to apply the model to an actual disease, we need to know the number removed per unit time, namely,  $dR/dt$  as a function of time. Using previous results, we can obtain an equation for  $R$  alone

$$\frac{dR}{dt} = aI = a(N_0 - R - S) = a\left(N_0 - R - S_0 \exp\left[-\frac{R}{\rho}\right]\right), \quad R(0) = 0, \quad (1)$$

which can only be solved in a parametric way. However this form is not convenient. Of course we can always compute the solution numerically if we know  $a$ ,  $b$ ,  $S_0$  and  $N_0$ . But, usually we don't know all the parameters. Thus, we try to carry out a best fit procedure, assuming, of course, that the model actually is a reasonable description of the epidemic.

Kermack and McKendrick argued that if the epidemic is not large,  $R/\rho$  is small. Using this observation, we can approximate equation (1) as

$$\begin{aligned} \frac{dR}{dt} &\approx a\left(N_0 - R - S_0\left[1 - \frac{R}{\rho} + \frac{1}{2}\left(\frac{R}{\rho}\right)^2\right]\right) \\ &= a\left(N_0 - S_0 + \left[\frac{S_0}{\rho} - 1\right]R - \left[\frac{S_0 R^2}{2\rho^2}\right]\right). \end{aligned} \quad (2)$$

Factoring the right-hand side quadratic in  $R$ , we can integrate the equation to get

$$\begin{aligned} R(t) &= \frac{\rho^2}{S_0} \left[ \left( \frac{S_0}{\rho} - 1 \right) + \alpha \tanh \left( \frac{\alpha t}{2} - \phi \right) \right] \\ \alpha &= \left[ \left( \frac{S_0}{\rho} - 1 \right)^2 + \frac{2S_0(N - S_0)}{\rho^2} \right]^{1/2}, \quad \phi = \frac{\tanh^{-1} \left( \frac{S_0}{\rho} - 1 \right)}{\alpha}. \end{aligned} \quad (3)$$

The removal rate is then given by

$$\frac{dR}{dt} = \frac{a\alpha^2\rho^2}{2S_0} \operatorname{sech}^2\left(\frac{\alpha at}{2} - \phi\right), \quad (4)$$

which involves only three parameters, namely  $a\alpha^2\rho^2/(2S_0)$ ,  $\alpha b$  and  $\phi$ . With epidemics that are not large, it is this function of time which we should fit to the Public Health records. On the other hand, if the disease is such that we know the actual number of the removed class, then it is  $R(t)$  that we should use. If  $R/\rho$  is not small, however, we must use the original differential equation for  $dR/dt$ .

## Examples

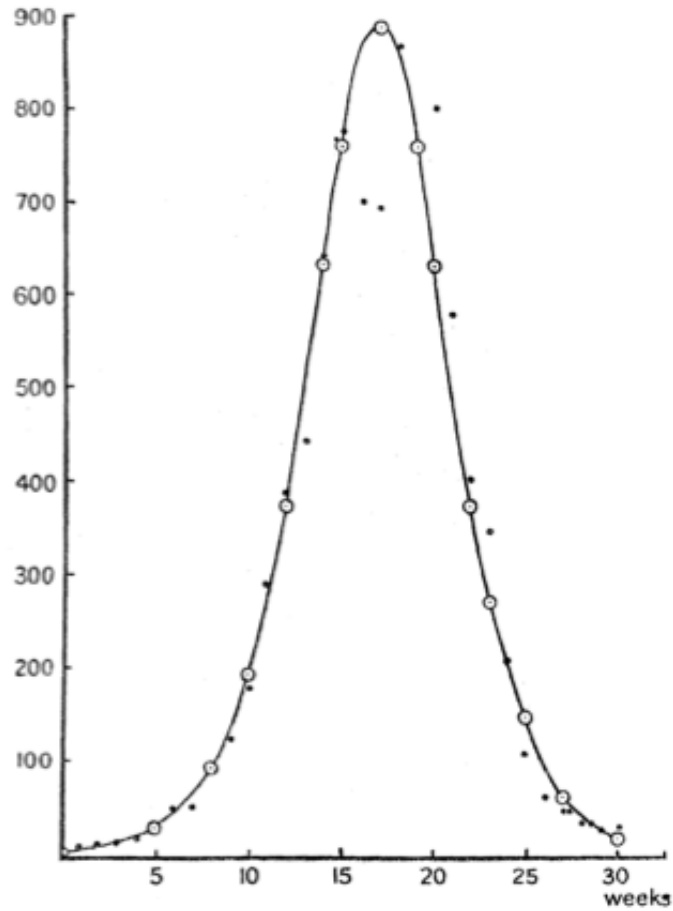
### Bombay Plague Epidemic, 1905-6

This epidemic lasted for almost a year. Most of the victims who got the disease died, the number removed per week, that is  $dR/dt$ , is approximately equal to the deaths per week. On the basis that the epidemic was not severe (relative to the population size), Kermack and McKendrick compared the actual data with (4) and determined the best fit for the three parameters and got

$$\frac{dR}{dt} = 890 \operatorname{sech}^2(0.2t - 3.4). \quad (5)$$

Figure 1 is from [4] showing the comparison between data and their model. Kermack and McKendrick [4] note

plague in man is a reflection of plague in rats, and that with respect to the rat (1) the uninfected population was uniformly susceptible; (2) that all susceptible rats in the island had an equal chance of being infected; (3) that the infectivity, recovery, and death rates were of constant value throughout the course of sickness of each rat; (4) that all cases ended fatally or became immune; and (5) that the flea population was so large that the condition approximated to one of contact infection. None of these assumptions are strictly fulfilled and consequently the numerical equation can only be a very rough approximation. A close fit is not to be expected, and deductions as to the actual values of the various constants should not be drawn. It may be said, however, that the calculated curve, which implies that the rates did not vary during the period of the epidemic, conforms roughly to the observed figures.



The accompanying chart is based upon figures of deaths from plague in the island of Bombay over the period December 17, 1905, to July 21, 1906. The ordinate represents the number of deaths per week, and the abscissa denotes the time in weeks. As at least 80 to 90 per cent. of the cases reported terminate fatally, the ordinate may be taken as approximately representing  $dz/dt$  as a function of  $t$ . The calculated curve is drawn from the formula

$$\frac{dz}{dt} = 890 \operatorname{sech}^2(0.2t - 3.4).$$

Figure 1: Bombay plague data and fit from [4]

We can also see the threshold effect in this model by examining the removed population (3). At the end of the epidemic ( $t \rightarrow \infty$ ), the removed population is

$$R(\infty) = \frac{\rho^2}{S_0} \left[ \left( \frac{S_0}{\rho} - 1 \right) + \alpha \right], \quad \alpha = \left[ \left( \frac{S_0}{\rho} - 1 \right)^2 + \frac{2S_0(N_0 - S_0)}{\rho^2} \right]^{1/2}.$$

Since  $N_0 - S_0$  is the initial number of infected individuals, it is reasonable to assume this number is small compared with  $S_0$  and we obtain

$$R(\infty) \approx \frac{2\rho}{S_0} \left[ \left( S_0 - \rho \right) \right].$$

If  $I_0$  is small then  $S_0 \approx N_0$ . If  $N_0$  is equal to (or less than)  $\rho$ , no epidemic can take place. If however,  $N_0$  slightly exceeds the value  $\rho$  then a small epidemic occurs. If we write  $N_0 = \rho + \epsilon$  then  $R(\infty) \approx 2\epsilon$ . To a first approximation, the size of the epidemic will be twice the excess if  $\epsilon$  is small compared with  $N_0$ . So at the end of the epidemic, the population will be just as far below the threshold density, as it initially was above it.

## Influenza Epidemic in an English Boarding School, 1978

In 1978, anonymous authors sent a note to the British Medical Journal reporting an influenza outbreak in a boarding school in the north of England. Figure 2 shows the data accompanying the note. Table 1 gives values, estimated from the figure, of the number of individuals confined to bed each day.

The outbreak was in a boys school with a total of 763 boys. Of these, 512 were confined to bed during the epidemic which lasted from the 22nd of January to the 4th of February. It also seems that one infected boy started the epidemic. Many of our model assumptions apply to this scenario; however, the epidemic is severe so we cannot use the approximation we made in the last example. Parameter fitting has to be done by solving the full ordinary differential equations of the SIR model.

We can take a simpler approach to get an estimate of the parameters describing this disease. The epidemic started with one sick boy, with two more getting sick one day later. Thus, we have  $I_0 = 1$ ,  $S_0 = 762$  and

$$\begin{aligned} \frac{dS}{dt} &\approx -2 \text{ per individual per day.} \\ a = \frac{-dS/dt}{SI} &\approx \frac{2}{762 \times 1} = 0.0026. \end{aligned}$$

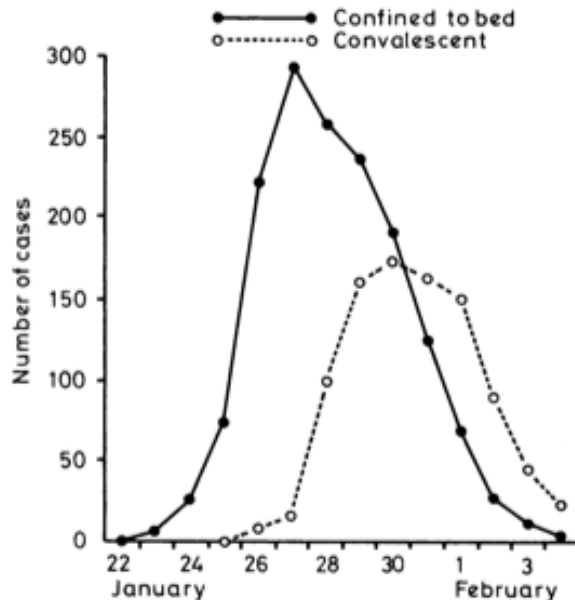


Figure 2: Influenza epidemic in an English Boarding School.

Now,  $b$  is the rate at which infected people are removed from the population. The report says that boys were taken to the infirmary within 1 or 2 days of becoming sick. So crudely, about  $1/2$  of the infected population was removed each day, or  $b = 0.5$  per day. This gives a value of  $\rho = b/a = 192$ . Figure 3 shows a plot of the phase plane with direction field and plots of  $S(t)$  and  $I(t)$  for these parameters. We predict  $I_{\max} = 306$  and  $S(\infty) = 16$  from the model.

Murray [3] reports performing a careful fit of model parameters using the full ODE model to obtain  $\rho = 202$ ,  $a = 2.18 \times 10^{-3}/\text{day}$ . The initial conditions are the same,  $N_0 = 763$ ,  $S_0 = 762$  and  $I_0 = 1$ . We note that these parameter values are close to our crude estimate and predict a similar course for the disease. The conditions for an epidemic are clearly met according to the model since  $S_0 > \rho$ . The epidemic is also severe since  $R/\rho$  is not always small;  $R(\infty) = 747$  using our ‘crude’ parameters (see Fig. 4).

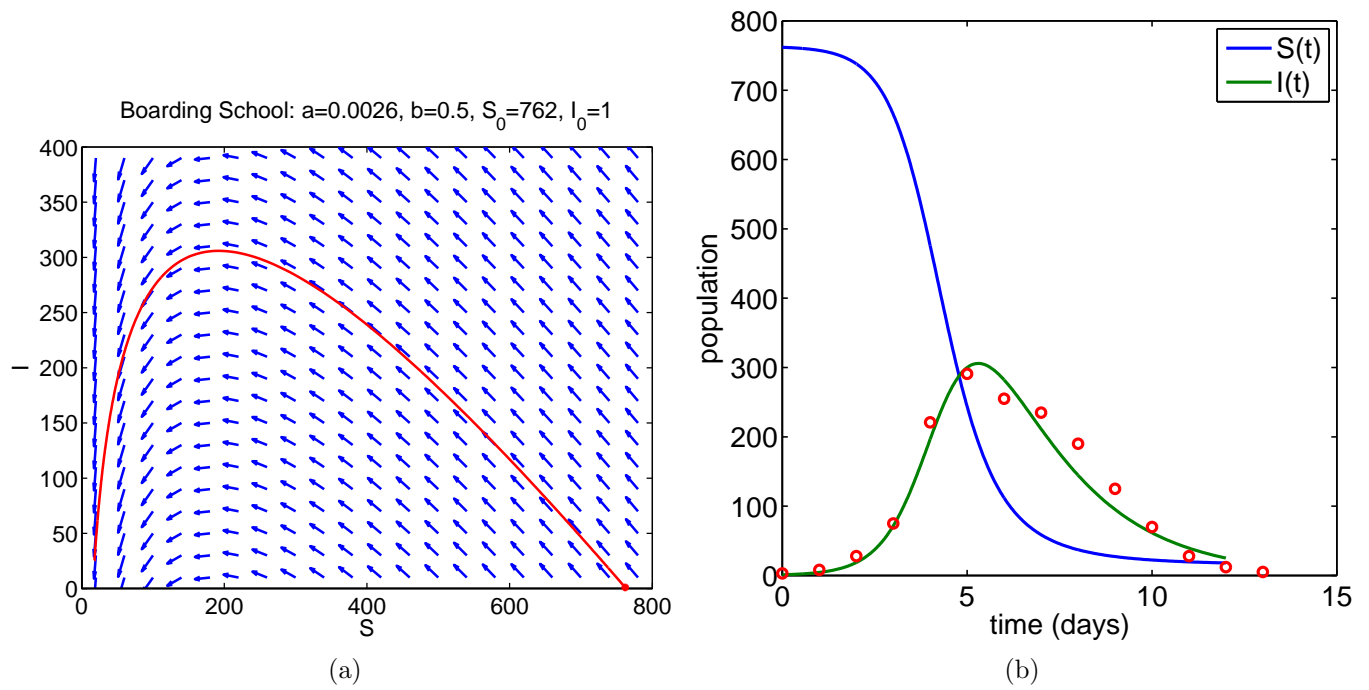


Figure 3: Model predictions compared to data for the English boys school using the SIR model with  $a = 0.0026$ ,  $b = 0.5$ ,  $S_0 = 762$  and  $I_0 = 1$ .

Day	Number
1	3
2	8
3	28
4	75
5	221
6	291
7	255
8	235
9	190
10	125
11	70
12	28
13	12
14	5

Table 1: Boarding School Data for Individuals Confined to Bed

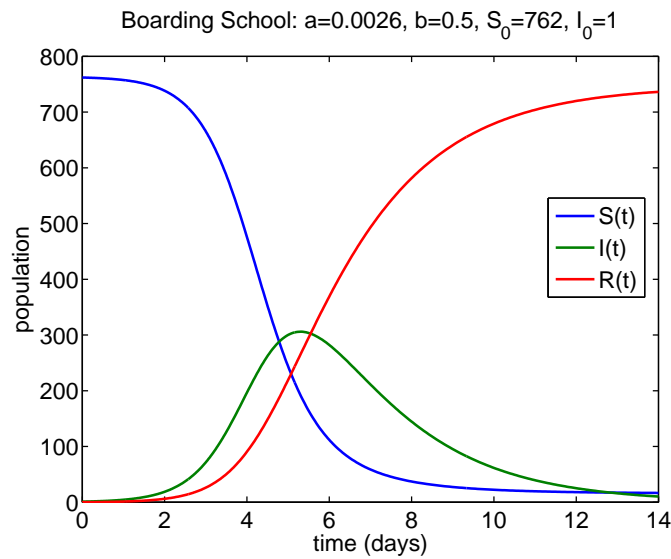


Figure 4: Model predictions compared to data for the English boys school using the SIR model with  $a = 0.0026$ ,  $b = 0.5$ ,  $S_0 = 762$  and  $I_0 = 1$ .

## Plague Outbreak in Eyam Village, 1665-66

In a wonderfully altruistic incident, the village of Eyam, England sealed itself off in 1665-66 when the plague was discovered, so as to prevent it spreading to neighboring villages. There is a museum in Eyam that tells the story (<http://www.eyammuseum.demon.co.uk/>). The villagers were successful in controlling the spread to other villages, but by the end of the epidemic only 83 of the original population of 350 survived. So we know  $S_0 = 350$  and  $S(\infty) = 83$ , but how do we obtain other information about parameters for a model? The discussion in this section is based heavily on [6].

The source of the plague in Eyam is attributed to the Great Plague of London (1664-1666) in which one sixth of the population succumbed to the disease. A tailor in Eyam received cloth from London which was infected with plague-carrying rat fleas that can produce plague in humans by biting their victims. The first victim was his assistant, George Viccars, who was buried September 7, 1665. Thereafter, the plague started to infect other victims as shown in Table 2. At this point the plague

1665	September	6 deaths
	October	23 deaths
	November	7 deaths
	December	9 deaths
1666	January	5 deaths
	February	8 deaths
	March	6 deaths
	April	9 deaths
	May	4 deaths

Table 2

appeared to be diminishing. But at the start of summer, the plague re-established itself. The next 5 months of deaths are shown in Table 3. Even though the dates and numbers are reported as deaths, they were actually burial dates. It is believed that burials were almost immediate for health reasons. However, toward the end of the outbreak there were no family members to bury the dead and burials might have been delayed. Marshall Howe was known to be a self-appointed grave digger who paid himself from the belongings of the dead.

The original strain of the Eyam plague is believed to be bubonic - from the rat fleas. It is believed that the outbreak did not continue in this form but turned at least partially to pneumonic, i.e. infected directly from person to person. This belief



1666	June	19 deaths
	July	56 deaths
	August	77 deaths
	September	24 deaths
	October	14 deaths

Table 3

is based on the fact that even though infected rodent fleas are temporarily important in the causation of bubonic plague in man, the continued existence of this form of the disease depends on the persistence of infected rodents. In this case, the original rodent was 150 miles away.

The SIR model is reasonable for this plague epidemic for the following reasons

1. The transmission of the plague is a rapidly spreading infectious disease.
2. The complete isolation of the village keeps  $N$  fixed. (This assumption is really only approximate since some wealthy villagers and some children fled. A few births and natural deaths were also recorded. However, the total population for the purposes of the model can be estimated at the outbreak time as the sum of final survivors and those who died during the plague. These figures are known.)
3. Only three cases of infected individuals are known to have survived. (One of these was Marshall Howe who presumably built up some immunity by being involved with repeated burials.) For the purpose of the model we can assume infected people were removed from the population by death. The time-dependent removed population is measurable from the list of dead.

The SIR model predicts a single peak  $I_{\max}$  given by

$$I_{\max} = N_0 - \rho + \rho \ln \frac{\rho}{S_0}. \quad (6)$$

However, the Eyam data contain at least two local infection peaks. We noticed the milder outbreak initially followed by more devastating later effects. Raggett [6] argues we can apply the SIR model starting in May or June, 1666. Then there is one peak in the data and the majority of the deaths are included. In the reference, Raggett starts on June 18. He also uses a time measurement of 15 1/2 days. The list of dead provides data for the deceased and removed (cumulative deceased). Table 4

Period (1666)	Deceased	Removed (measured at end of period)
June 19 - July 3/4	11.5	11.5
July 4/5 - July 19	26.5	38
July 20 - Aug 3/4	40.5	78.5
Aug 4/5 - Aug 19	41.5	120
Aug 20 - Sept 3/4	25	145
Sept 4/5 - Sept 19	11	156
Sept 20 - Oct 4/5	11.5	167.5
Oct 5/6 - Oct 20	10.5	178

Table 4: Deceased and Removed Populations over the Major Outbreak Period.

gives the data averages because of the half-day in each time interval. The initial time is taken as June 18, 1666 when  $R(0) = 0$  and  $N_0 = 261$ . This last figure is obtained by subtracting the 89 prior deaths from the initial population of 350. (Twelve of the 19 deaths recorded in June occurred prior to June 19.)

Note that the incubation period for human plague is a maximum of 6 days and the length of the illness is  $5 \frac{1}{2}$  days. Let's use a uniform period of 11 days for the total infection period, as in Raggett. Thus as the end of each time interval, one measures the infective population by examining the death register for the following 11 days. Using the information so obtained for the removed and infective populations, the susceptible population is directly estimated at the end of each period using  $N_0 = S(t) + I(t) + R(t)$ . Table 5 gives these population estimates. Given these populations, can we determine values of  $a$  and  $b$  that describe the plague outbreak of Eyam?

Using the equation

$$S(\infty) = S_0 \exp[-R(\infty)/\rho] = S_0 \exp(-(N_0 - S(\infty))/\rho), \quad (7)$$

from yesterday, we can estimate  $\rho \approx 159$  using  $S(\infty) = 83$ , the number of surviving villagers. Furthermore, using this value of  $\rho$  gives  $I_{\max} \approx 27$  with a corresponding number of susceptibles being  $\rho$  when the number of infectives is  $I_{\max}$ . Thus the corresponding number of removed is 75. From the table of data, the removed population is 70 and 77 on Aug 2 and 3; respectively. So, we have an estimate of the peak infection time. (Taking  $I$  to be 27 on both these dates gives a prediction of infection periods as 11 days in each case; consistent with our initial assumption.) Thus, we have an estimate for  $\rho = b/a$ . How can we estimate  $a$  and  $b$  individually?

Date (1666)	$S(t)$	$I(t)$
July 3/4	235	14.5
July 19	201	22
Aug 3/4	153.5	29
Aug 19	121	20
Sept 3/4	108	8
Sept 19	97	8
Oct 4/5	unknown	unknown
Oct 20	83	0

Table 5: Susceptible and Infective Populations at Terminal Period Dates.  $S(0) = 254$ ,  $I(0) = 7$ ,  $R(0) = 0$ ,  $N_0 = 261$

We assumed an 11 day infection period, so we would expect a removal rate of unity over 11 days. Using linear arguments gives an estimated removal rate of about 2.82 based on a 31 day period. The solution to the SIR model using these estimates is shown in Fig. 5.

If we numerically solve the SIR model using  $S(0) = 254$ ,  $I(0) = 7$  and  $R(0) = 0$  and  $\rho = 159$ , varying  $b$ , then setting  $a = b/159$  we can try to find a best fit to the data in Table 5. We would solve the equations from  $t = 0$  to  $t = 4$  and output 8 data values at time steps of 0.5. Each of these 8 population values are compared to the data in the table. In this way, we find  $b = 2.78$ . We need to remember that the computed time step of one unit corresponds to a real time interval of 31 days. The value 2.78 obtained from this fit is quite close to the estimate 2.82 given above.

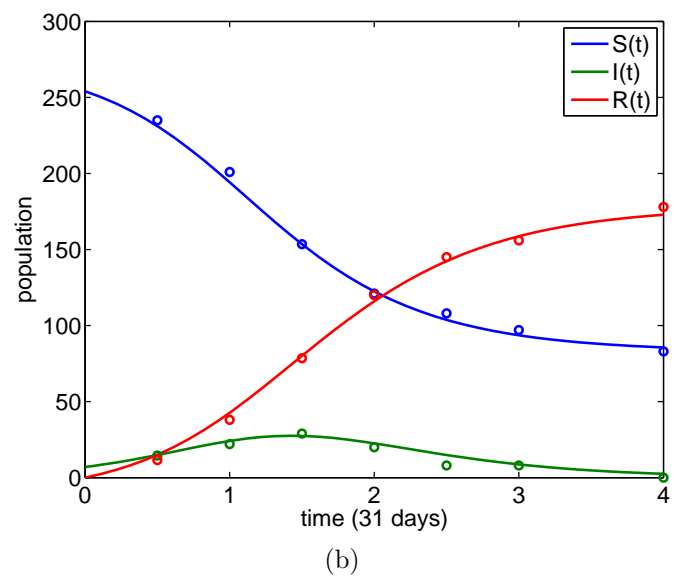
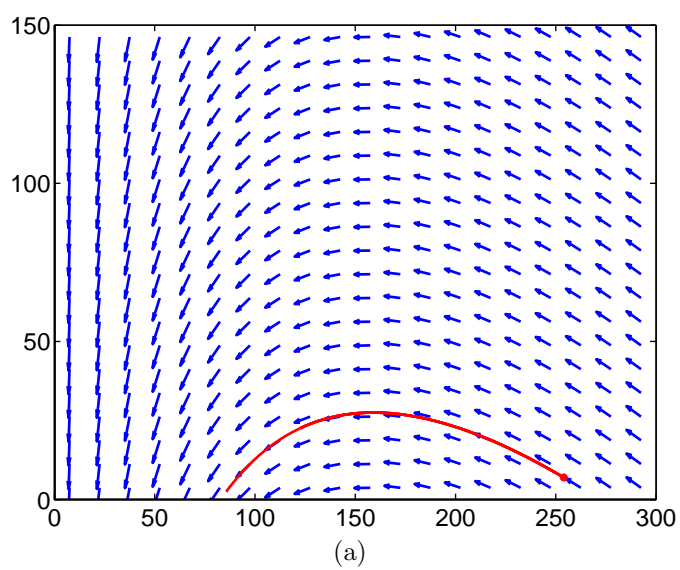


Figure 5: Model predictions compared to data for the Eyam plague outbreak using the SIR model with  $b = 2.82$ ,  $a = b/159$ ,  $S_0 = 254$ ,  $I_0 = 7$  and  $R(0) = 0$ .

Things to try:

1. Carry out the least-squares fit to the data for the Bombay plague epidemic. Try to determine parameters for the SIR model and numerically solve the model equations.
2. Carry out the least-squares fit to the data for the Eyam plague to find the parameter  $b$ .
3. Consider modeling the epidemics in this section using a discrete-time model. What are appropriate parameters for the model? How do these parameters compare with the continuous time model?

## References

- [1] Keshet, Leah, "Mathematical Models in Biology." Random House, 1988.
- [2] Hoppenstadt, F.C. and C.S. Peskin, "Mathematics in Medicine and the Life Sciences." Springer-Verlag, 1992.
- [3] Murray, J.D., "Mathematical Biology." Springer-Verlag, 1989.
- [4] Kermack, W.O. and A.G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics." Proceedings of the Royal Society of London. Series A, 115:700-721, 1927.
- [5] Abbey, H, "An examination of the Reed Frost theory of epidemics." Human Biology, 24:201-233, 1952.
- [6] Raggett, G.F., "Modeling the Eyam Plague," Bulletin of the Institute of Math. and its Appls., 18:221-226, 1982.